METHOD FOR THE DISSOCIATION OF THE EXTRACELLULAR HAEMOGLOBIN MOLECULE OF *ARENICOLA MARINA*, CHARACTERIZATION OF THE PROTEIN CHAINS CONSTITUTING SAID MOLECULE AND THE NUCLEOTIDE SEQUENCES ENCODING SAID PROTEIN CHAINS

A subject of the present invention is a method for the dissociation of the extracellular haemoglobin molecule of Annelida, in particular of *Arenicola marina*, as well as the characterization of the protein chains constituting said molecule.

A subject of the present invention is also the characterization of the nucleotide sequences encoding the abovementioned protein chains, as well as the method for preparing these nucleotide sequences.

Blood is a complex liquid the main function of which is to transport oxygen and carbon dioxide, in order to ensure the respiratory processes. It is the haemoglobin molecule, which is found in the red blood cells, which ensures this function.

The haemoglobin molecule in mammals is formed by an assembly of four similar functional polypeptide chains in pairs (2 chains of type α globin and 2 chains of type β globin). Each of these polypeptide chains possesses the same tertiary structure of a myoglobin molecule (Dickerson and Geis, 1983).

Heme, the active site of haemoglobin, is a tetrapyrrolic protoporphyrin ring, containing a single iron atom in its centre. The iron atom, which fixes oxygen, establishes 6 coordinancy bonds: four with the nitrogen atoms of the porphyrin, one with the proximal histidine F8 and one with the oxygen molecule during the oxygenation of the globin.

We are currently faced with blood supply problems, due to the reduction in blood donations for fear of contamination. Thus, research into blood substitutes has accelerated over the last few years. We are seeking to design artificial blood substitutes capable of eliminating the risks of transmission of infectious diseases, but also to become free from problems relating to blood group compatibility.

Up to now, the main research routes relate to the synthesis of chemical products on the one hand (Clark and Gollan, 1966) and the synthesis of biological products on the other hand (Chang, 1957; Chang, 1964).

As regards the first research route, perfluorocarbons (PFCs) have been used. The PFCs are chemical products capable of transporting oxygen and they can dissolve a large quantity of gas, such as oxygen and carbon dioxide.

At present, attempts are being made to produce emulsions of these products which could be dispersed in the blood more effectively (Reiss, 1991; Reiss, 1994; Goodin et al., 1994).

The advantage of the PFCs resides in their oxyphoric capacity which is directly proportional to the quantity of oxygen to be found in the lungs. Moreover, because of the absence of a membrane to pass through, the PFCs can transport oxygen more rapidly towards the tissues. However, the long-term effects of retention of these products in the organism are not known. When these products were used for the first time in the 1960s as a blood substitute in mice (Clark and Gollan, 1966; Geyer et al., 1966; Sloviter and Kamimoto, 1967), the side effects were very significant. The PFCs were not eliminated from the circulation in a satisfactory manner and accumulated in the tissues of the organism, which caused œdema.

In the 1980s, a new version PFC was tested in the clinical phase. But the problems of storage, financial cost, non-negligible side effects and the low effectiveness of this compound prevented the extension of its marketing (Naito, 1978; Mitsuno and Naito, 1979; Mitsuno and Ohyanagi, 1985).

Recently, a new generation of PFC (PFBO perfluorooctylbromide) has been developed. A novel product (Reiss, 1991) is undergoing clinical trials in the United States, but it has already been noted that an increase in the quantity of oxygen in the blood can create an accumulation of oxygen in the tissues, which is dangerous to the organism (formation of superoxide-type oxygen radical).

Thus, in spite of the gradual progress made, the side effects of these compounds are still too significant for them to be marketed on a large scale.

As regards the second research route, researchers have worked on the development of blood substitutes by modifying the structure of natural haemoglobin (Chang, 1957; Chang, 1997). In order to obtain a blood substitute of modified

haemoglobin type, haemoglobins synthesized by genetically modified microorganisms, or of human or animal origin are used, in particular the molecule of bovine haemoglobin. In fact, bovine haemoglobin is slightly immunologically different from human haemoglobin, but it transports oxygen towards the tissues more easily. Nevertheless, the risk of viral contamination or contamination of spongiform encephalopathy type still remain significant.

In order to be functional, the haemoglobin must be in contact with an allosteric effector, 2,3-diphosphoglycerate (2,3-DPG), present only inside the red blood cells (Dickerson and Geis, 1983). Moreover, without 2,3-DPG and other elements present in the red blood cells such as methaemoglobin reductase, haemoglobin undergoes an auto-oxidation process and loses its ability to transport oxygen or carbon dioxide.

These processes can be eliminated by modifying the structure of the haemoglobin, and more precisely by stabilizing the weak bonds of the tetrameric molecule between the two α and β dimers (Chang, 1971). Numerous modifications have been tested: covalent bond between two α chains, between two β chains or also between α and β (Payne, 1973; Bunn and Jandl, 1968).

Attempts have also been made to polymerize the tetrameric molecules or to conjugate them with a polymer named polyethylene glycol (PEG) (Nho et al., 1992). These modifications have the consequence of stabilizing the molecule and increasing its size, preventing its elimination by the kidneys.

The Annelida have been much studied for their extracellular haemoglobin (Terwilliger, 1992; Lamy et al., 1996). These extracellular haemoglobin molecules are present in the three classes of Annelida: Polychaetes, Oligochaetes and Achaetes and even in the Vestimentifera. These are giant biopolymers, made up of approximately 200 polypeptide chains belonging to 6 or 8 different types which are generally divided into two categories. The first category, comprising 144 to 192 elements, includes the so-called "functional" polypeptide chains carrying an active site and capable of reversibly binding oxygen; these are globin-type chains the masses of which are comprised between 15 and 18 kDa and which are very similar to the α and β type chains of vertebrates. The second category, comprising 36 to 42 elements, includes the so-called "structural" polypeptide chains possessing few or no active sites but allowing the assembly of "twelfths".

The first images of extracellular haemoglobins of Arenicola obtained (Roche et al., 1960) revealed hexagonal structures. Each haemoglobin molecule is made up of two superimposed hexagons (Levin, 1963; Roche, 1965) described as a "hexagonal bilayer" and each hexagon is itself formed by the assembly of six elements in the form of a drop of water (Van Bruggen and Weber, 1974; Kapp and Crewe, 1984), described as a "hollow globular structure" (De Haas et al., 1996) or "twelfth". The native molecule is formed of twelve of these sub-units, with a molecular mass of approximately 250 kDa.

Thus, the French patent no. 2 809 624 relates to the use as a blood substitute of extracellular haemoglobin of *Arenicola marina*, a Polychaete Annelida of the intertidal ecosystem, said blood substitute making it possible to eliminate the problems of a shortage of donations.

Although the overall architecture of the haemoglobin of *Arenicola marina* is known, in particular thanks to its fine quaternary study by mass spectrometry (Zal et al., 1997), the primary sequences of the different protein chains which compose it are not.

Thus, the purpose of the present invention is to provide the protein sequences which compose the haemoglobin molecule of *Arenicola marina*.

Another purpose of the present invention is to provide the first stages of *in vitro* synthesis of extracellular haemoglobin of *Arenicola marina* in order to develop a blood substitute by means of biochemistry and molecular biology methods.

Another purpose of the present invention is to provide a method for preparing the haemoglobin molecule, optionally simplified, by genetic engineering, in order in particular to increase the stock of this molecule within the framework of use as a blood substitute.

The present invention relates to a method for the dissociation of the extracellular haemoglobin molecule of Annelida, in particular of *Arenicola marina*, making it possible to obtain protein chains constituting said molecule,

said method being characterized in that it comprises a stage of bringing together a sample of extracellular haemoglobin of Annelida, in particular of *Arenicola marina* and at least one dissociating agent, in particular a mixture made up of dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine hydrochloride (TCEP) or beta-mercaptoethanol and a dissociation buffer, for a sufficient time to separate the protein chains from each other.

The present invention relates to a method for obtaining protein chains constituting the extracellular haemoglobin molecule of Annelida, in particular of *Arenicola marina*,

said method being characterized in that it comprises a stage of bringing together a sample of extracellular haemoglobin of Annelida, in particular of *Arenicola marina* and at least one dissociating agent, and if appropriate a reducing agent, in particular a mixture made up of dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine hydrochloride (TCEP) or beta-mercaptoethanol and a dissociation buffer, for a sufficient time to separate the protein chains from each other.

The term "extracellular haemoglobin" designates a haemoglobin not contained in the cells and dissolved in the blood.

The expression "dissociation" designates a chemical treatment capable of breaking weak interactions (hydrophobic, electrostatic, hydrogen etc.).

The term "dissociation buffer" designates a liquid containing a buffer making it possible to adjust the pH and containing dissociating agents.

The expression "dissociating agent" designates a chemical compound capable of breaking weak interactions (hydrophobic, electrostatic, hydrogen etc.). Said dissociating agent is in particular chosen from: hydroxide ions, urea or heteropolytungstate ions or guanidinium salts or SDS (sodium dodecyl sulphate).

The expression "reduction" designates a chemical treatment capable of breaking strong interactions such as disulphide bridges.

The expression "reducing agent" designates a chemical compound capable of breaking strong interactions such as disulphide bridges.

The ten protein chains constituting the extracellular haemoglobin molecule of *Arenicola marina* include 8 globin-type chains and 2 structural-type chains.

It is recalled that the extracellular haemoglobin of *Arenicola marina* with a mass of 3648 ± 24 kDa is made up of 198 polypeptide chains belonging to 10 different types divided into two categories:

– the first (156 chains) includes so-called "functional" polypeptide chains carrying an active site capable of reversibly binding oxygen; these are globin-type chains the masses of which are comprised between 15 and 18 kDa; these chains are very similar to the α and β-type chains of vertebrates; and

– the second (42 chains) includes so-called "structural" (or "linker") polypeptide chains possessing few or no active sites but allowing the assembly of the dodecamers; these chains have molecular masses comprised between 22 and 27 kDa.

The present invention relates to a method for the dissociation of the extracellular haemoglobin molecule of Arenicola marina, making it possible to obtain protein chains constituting said molecule,

said method being characterized in that it comprises a stage of bringing together a sample of extracellular haemoglobin of Arenicola marina and a mixture made up of dithiothreitol (DTT) and a dissociation buffer, for approximately one hour to three weeks.

An advantageous dissociation method of the invention is characterized in that the dissociation buffer comprises a buffering agent at a concentration comprised between approximately 0.05 M and approximately 0.1 M, in particular Trisma (tris[hydroxymethyl]aminomethane), hepes, sodium phosphate, sodium borate, ammonium bicarbonate or ammonium acetate, and 0 to 10 mM of EDTA adjusted to a pH comprised between approximately 5 and approximately 12, and preferably between approximately 7.5 and 12, the whole being in particular adjusted to a pH comprised between approximately 2 and 12, and preferably between approximately 5 and 12.

Preferably, said dissociation buffer comprises EDTA at a concentration of approximately 1 mM adjusted to a pH of approximately 10, in particular with a 2N solution of soda.

According to an advantageous embodiment, the method of the invention is characterized in that the protein chains constituting said molecule are obtained by the reduction of four sub-units by a reducing agent, for example in the presence of DTT, said sub-units themselves being obtained by bringing together a sample of extracellular haemoglobin of *Arenicola marina* and different dissociating agents, in particular a dissociation buffer.

The native molecule is dissociated into sub-units under the action of non-reducing dissociating agents. There is therefore no breakage of the covalent bonds. However, after the action of a reducing agent (cleavage of the covalent bonds), the sub-units are reduced to polypeptide chains (protein chains made up of the assembly of amino acids).

The abovementioned 4 sub-units are therefore: monomers, dimers, trimers and dodecamers.

The monomers are globin chains.

The dimers in the homo form and heterodimers are structural chains.

The trimers are covalent assemblies of three globin chains.

The dodecamers are made up of 12 protein chains; for example: 3 trimers + 3 monomers, 2 trimers + 6 monomers, 1 trimer + 9 monomers.

It is therefore possible to obtain the protein chains either in a single stage by direct reduction of the extracellular haemoglobin of *Arenicola marina*, or in two stages, one consisting of the dissociation of the extracellular haemoglobin of *Arenicola marina* into 4 sub-units and the other being the reduction of said 4 sub-units into protein chains.

The present invention also relates to a dissociation method as defined above, characterized in that the dissociating agents used in order to obtain the abovementioned 4 sub-units are the following:

- a dissociation buffer solution comprising: 0.1 M of Trisma base (tris[hydroxymethyl]aminomethane) and 0 to 10 mM of EDTA adjusted to a pH comprised between approximately 5 and approximately 12, and preferably between approximately 7.5 and approximately 12, and
- a urea solution, the concentration of which is comprised between approximately 0.1 M and approximately 8 M, and is in particular equal to 3 M.

The present invention also relates to a dissociation method as defined above, characterized in that the dissociating agents for obtaining the 4 sub-units are the following:

- a dissociation buffer solution comprising 0.1 M of Trisma base and 1 mM of EDTA adjusted to pH 10, and
- 3 M of urea.

The present invention also relates to a dissociation and reduction method as defined above, characterized in that the dissociating and reducing agents used in order to obtain the protein chains are the following:

- a dissociation buffer solution comprising: 0.1 M of Trisma base (tris[hydroxymethyl]aminomethane) at a pH comprised between approximately 8 and approximately 9, and
- a urea solution, the concentration of which is comprised between approximately 4 M and approximately 8 M, and is in particular equal to 8 M, and
- 1 to 10% DTT

or

- a dissociation buffer solution comprising: 0.1 M of ammonium bicarbonate at a pH comprised between approximately 8 and approximately 9, and
- 1 to 10% DTT

The dissociation and reduction method of the invention makes it possible to obtain a composition containing the mixture of the protein chains constituting the extracellular haemoglobin molecule of *Arenicola marina*.

The present invention also relates to a method for preparing primer pairs from the protein chains as obtained according to the method as defined above, said method being characterized in that it comprises the following stages:

– the isolation of each of the protein chains constituting the haemoglobin molecule as obtained according to the method as defined above,

– the microsequencing of each of the abovementioned isolated protein chains by mass spectrometry and Edman sequencing, in order to obtain a microsequence corresponding to each of the sequences made up of 5 to 20 amino acids, and

– the determination of the degenerated primers from the abovementioned microsequences.

The first stage of isolation of the protein chains is in particular carried out by Reversed-phase liquid chromatography and two-dimensional gel from the abovementioned mixture comprising the protein chains constituting the haemoglobin molecule as obtained according to the dissociation and reduction method of the invention.

The expression "microsequence" designates fragments of protein sequences.

The abovementioned microsequences can originate both from the C- and N-terminal ends but also from internal sequences.

The protein chains can be obtained by Reversed-phase liquid chromatography or from 2D gel from purified haemoglobin of *Arenicola marina.* Each peak or spot was cut out and digested by a protease. The peptides thus obtained were extracted from the gels and separated by capLC (capillary liquid chromatography). The fragments are then analyzed by mass spectrometry. On the other hand, each peak isolated by Reversed-phase was analyzed by Edman sequencing.

The expression "degenerated primers" designates nucleotide sequences obtained from fragments of protein sequences. They are called degenerated primers because of the degeneration of the genetic code (several codons for 1 amino acid).

The last stage of determination of the degenerated primer pairs corresponds to their synthesis.

This stage makes it possible to obtain both sense primers and antisense primers.

The present invention also relates to primer pairs as obtained according to the method as defined above, said pairs being in particular the following:

a) *Sense primer*: GAR TGY GGN CCN TTR CAR CG (SEQ ID NO: 21)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)
b) *Sense primer*: TGY GGN ATH CTN CAR CG (SEQ ID NO: 23)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)
c) *Sense primer*: AAR GTI AAR CAN AAC TGG (SEQ ID NO: 24)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)
d) *Sense primer*: TGY TGY AGY ATH GAR GAY CG (SEQ ID NO: 25)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)
e) *Sense primer*: AAR GTN ATH TTY GGN AGR GA (SEQ ID NO: 26)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)
f) *Sense primer*: GAR CAY CAR TGY GGN GGN GA (SEQ ID NO: 27)
Antisense primer: CTC CTC TCC TCT CCT CTT CCT (SEQ ID NO: 22)

where:

R represents A or G,

Y represents C or T,

N represents A, G, C or T,

I represents inosine,

H represents A, C or T.

The present invention also relates to primer pairs as obtained according to the method as defined above, said pairs being in particular the following:

a) *Sense primer*: GAR TGY GGN CCN TTR CAR CG SEQ ID NO: 21
Antisense primer: CCA NGC NTC YTT RTC RAA GCA SEQ ID NO: 28
b) *Sense primer*: AN TGY GGN CCN CTN CAR CG SEQ ID NO: 29
Antisense primer: CCA NGC NTC YTT RTC RAA GCA SEQ ID NO: 28
c) *Sense primer*: AAR GTI AAR CAN AAC TGG SEQ ID NO: 24
Antisense primer: CCA NGC NCC DAT RTC RAA SEQ ID NO: 30
d) *Sense primer*: TGY TGY AGY ATH GAR GAY CG SEQ ID NO: 25
Antisense primer: CA NGC NYC RCT RTT RAA RCA SEQ ID NO: 31

where:

R represents A or G,

Y represents C or T,

N represents A, G, C or T,

I represents inosine,

D represents A, G or T,

H represents A, C or T.

The present invention also relates to a method for preparing nucleotide sequences encoding the protein chains constituting the haemoglobin molecule of *Arenicola marina*, from the primers as obtained according to the method as defined above, said method being characterized in that it corresponds to a polymerase chain amplification method (PCR), comprising a repetition of at least 30 times the cycle constituted by the following stages:

* the denaturation, by heating, of the monocatenary cDNA encoding one of the protein chains constituting the haemoglobin molecule of *Arenicola marina*, so as to denature any secondary structures and RNA residuals, said cDNA being obtained from mRNA, this stage making it possible to obtain strands of denatured monocatenary cDNA,

* the hybridization of the primer pairs as obtained by the method as defined above to the strands of abovementioned denatured monocatenary cDNA at an appropriate temperature, in order to obtain hybridized primers, and

* the synthesis of the complementary strand of the cDNA by a polymerase at an appropriate temperature, from the hybridized primers as obtained in the preceding stage.

The cDNA encoding the abovementioned protein chains is obtained from mRNA, said mRNA being obtained by purification from total RNAs extracted from growing juvenile Arenicolae, said juvenile Arenicolae having a high level of transcription of the different messenger RNAs.

If the abovementioned cycle is repeated less than 30 times, the amplification of the DNA is reduced.

The expression "optional secondary structures" designates anarchic pairings between two sequences of cDNA.

The expression "denaturation by heating of the cDNA" designates the breaking of the anarchic pairings between two sequences of cDNA.

The expression "hybridization at an appropriate temperature" designates the recognition by the primers of their complementary sequences on the DNA matrix.

According to an advantageous embodiment, the method for preparing nucleotide sequences according to the invention is characterized in that:

– the first stage of said method is a stage of denaturation of the cDNA encoding one of the protein chains constituting the haemoglobin molecule of *Arenicola marina* of approximately 10 seconds to approximately 5 minutes at a temperature comprised between approximately 90°C and approximately 110°C,

– the cycle, repeated approximately 30 to 40 times, comprises the following stages:

* a stage of denaturation of the cDNA encoding one of the protein chains constituting the haemoglobin molecule of *Arenicola marina* of approximately 10 seconds to approximately 5 minutes, at a temperature comprised between approximately 90°C and approximately 110°C,
* a stage of hybridization of the primer pairs of the invention to the abovementioned strands of monocatenary cDNA in order to obtain hybridized primers, of approximately 20 seconds to approximately 2 minutes, at a temperature comprised between approximately 50°C and approximately 60°C, preferably between approximately 50°C and approximately 56°C,
* a stage of elongation of the hybridized primers as obtained previously by a polymerase of approximately 20 seconds to approximately 1 minute and 30 seconds, at a temperature comprised between approximately 70°C and approximately 75°C, and

– the last stage of the method is a stage of elongation of the hybridized primers as obtained previously by a polymerase of approximately 5 minutes to approximately 15 minutes at a temperature comprised between approximately 70°C and approximately 75°C.

The PCR reaction of the method of the invention is in particular carried out in the presence of cDNA (5 to 20 ng), sense (100 ng) and antisense (100 ng) primer, dNTP (200 µM final), $MgCl_2$ (2 mM final), PCR buffer (supplied with the polymerase) (1 X final), Taq polymerase (1 unit) and water (25 µl qsf).

The method for preparing the abovementioned nucleotide sequences makes it possible to obtain partial coding sequences.

Once the partial coding sequences have been obtained by means of the preceding experiments, the amplification and the sequencing of the whole of the coding sequence

of the cDNA of globins and of the linker are carried out by 5' RACE (Rapid Amplification cDNA Ends) PCR and according to the protocol recommendations of the data sheet provided by the supplier (5'/3' RACE kit, Roche).

The present invention also relates to a preparation method as defined above, characterized in that the primer pairs used are as defined previously.

A particularly advantageous preparation method according to the invention is a method for preparing nucleotide sequences as defined above, characterized in that the pair of primers used is: (GAR TGY GGN CCN TTR CAR CG; CCA NGC NTC YTT RTC RAA GCA) or (GAR TGY GGN CCN TTR CAR CG; CTC CTC TCC TCT CCT CTT CCT), R, Y and N being as defined above,

said method being characterized in that:

- the first stage of the method is a stage of denaturation of the cDNA encoding the protein chains constituting the haemoglobin molecule of *Arenicola marina*, of 4 minutes at a temperature equal to 95°C,
- the cycle, repeated 35 times, comprises the following stages:
  * a stage of denaturation of the cDNA encoding one of the protein chains constituting the haemoglobin molecule of *Arenicola marina*, of 30 seconds at a temperature equal to 95°C,
  * a stage of hybridization of the primer pairs of the invention to the abovementioned strands of monocatenary cDNA in order to obtain hybridized primers, of 30 seconds at a temperature equal to 56°C,
  * a stage of elongation of the hybridized primers as obtained previously by a polymerase of 40 seconds at a temperature equal to 72°C, and
- the last stage of the method is a stage of elongation of the hybridized primers as obtained previously by a polymerase of 10 minutes at a temperature equal to 72°C,

in order to obtain the nucleotide sequence SEQ ID NO: 13,

said method optionally comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 1.

The partial sequence SEQ ID NO: 13 was then completed by 5' RACE PCR experiments as explained above. The nucleotide sequence SEQ ID NO: 13 is a novel nucleotide sequence encoding a protein chain corresponding to a globin chain, denoted A2a. SEQ ID NO: 13 comprises 376 nucleotides.

The nucleotide sequence SEQ ID NO: 1 (from the start codon to the stop codon, i.e. the transcribed and translated sequence which corresponds to a functional globin monomer) is a novel nucleotide sequence encoding a protein chain corresponding to the abovementioned globin chain, denoted A2a. SEQ ID NO: 1 comprises 474 nucleotides.

A particularly advantageous preparation method according to the invention is a method for preparing nucleotide sequences as defined above, characterized in that the pair of primers used is the following: (AN TGY GGN CCN CTN CAR CG; CCA NGC NTC YTT RTC RAA GCA), N, Y and R being as defined above,

said method being characterized in that:

- the first stage of the method is a stage of denaturation of the cDNA encoding the protein chains constituting the haemoglobin molecule of *Arenicola marina*, of 4 minutes at a temperature equal to 95°C,
- the cycle, repeated 35 times, comprises the following stages:
  * a stage of denaturation of the cDNA encoding one of the protein chains constituting the haemoglobin molecule of *Arenicola marina*, of 30 seconds at a temperature equal to 95°C,
  * a stage of hybridization of the primer pairs of the invention to the abovementioned strands of monocatenary cDNA in order to obtain hybridized primers, of 30 seconds at a temperature equal to 52°C,
  * a stage of elongation of the hybridized primers as obtained previously by a polymerase of 40 seconds at a temperature equal to 72°C, and
- the last stage of the method is a stage of elongation of the hybridized primers as obtained previously by a polymerase of 10 minutes at a temperature equal to 72°C,

in order to obtain the nucleotide sequence SEQ ID NO: 15.

The nucleotide sequence SEQ ID NO: 15 is a novel nucleotide sequence encoding a protein chain corresponding to a globin chain, denoted A2b. SEQ ID NO: 15 comprises 288 nucleotides.

A particularly advantageous preparation method according to the invention is a method for preparing nucleotide sequences as defined above, characterized in that the pair of primers used is the following: (TGY GGN ATH CTN CAR CG; CTC CTC TCC TCT CCT CTT CCT), N, Y and R being as defined above,

said method being characterized in that:

– the first stage of the method is a stage of denaturation of 4 minutes at a temperature equal to 95°C,

– the cycle, repeated 35 times, comprises the following stages:

* a stage of denaturation of 30 seconds at a temperature equal to 95°C,
* a stage of hybridization of 30 seconds at a temperature equal to 53°C,
* a stage of elongation of 40 seconds at a temperature equal to 72°C, and

– the last stage of the method is a stage of elongation of 10 minutes at a temperature equal to 72°C,

in order to obtain the nucleotide sequence SEQ ID NO: 15,

said method optionally comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 3.

The partial sequence SEQ ID NO: 15 was then completed by 5' RACE PCR experiments as explained above.

The nucleotide sequence SEQ ID NO: 3 (from the start codon to the stop codon, i.e. the transcribed and translated sequence which corresponds to a functional globin monomer) is a novel nucleotide sequence encoding a protein chain corresponding to the abovementioned globin chain, denoted A2b. SEQ ID NO: 3 comprises 477 nucleotides.

A particularly advantageous preparation method according to the invention is a method for preparing nucleotide sequences as defined above, characterized in that the pair of primers used is: (AAR GTI AAR CAN AAC TGG; CCA NGC NCC DAT RTC RAA) or (AAR GTI AAR CAN AAC TGG; CTC CTC TCC TCT CCT CTT CCT), R, I, N and D being as defined above,

said method being characterized in that:

– the first stage of the method is a stage of denaturation of the cDNA encoding each of the protein chains constituting the haemoglobin molecule of *Arenicola marina*, of 4 minutes at a temperature equal to 95°C,

– the cycle, repeated 35 times, comprises the following stages:

* a stage of denaturation of the cDNA encoding one of the protein chains constituting the haemoglobin molecule of *Arenicola marina*, of 1 minute at a temperature equal to 95°C,
* a stage of hybridization of the primer pairs of the invention to the abovementioned strands of monocatenary cDNA in order to obtain hybridized primers, of 1 minute at a temperature equal to 50°C,

* a stage of elongation of the hybridized primers as obtained previously by a polymerase of 1 minute and 30 seconds at a temperature equal to 72°C, and

– the last stage of the method is a stage of elongation of the hybridized primers as obtained previously by a polymerase of 10 minutes at a temperature equal to 72°C,

in order to obtain the nucleotide sequence SEQ ID NO: 17,

said method optionally comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 5.

The partial sequence SEQ ID NO: 17 was then completed by 5' RACE PCR experiments as explained above. The nucleotide sequence SEQ ID NO: 17 is a novel nucleotide sequence encoding a protein chain corresponding to a globin chain, denoted A1. SEQ ID NO: 17 comprises 360 nucleotides.

The nucleotide sequence SEQ ID NO: 5 (from the start codon to the stop codon, i.e. the transcribed and translated sequence which corresponds to a functional globin monomer) is a novel nucleotide sequence encoding a protein chain corresponding to the abovementioned globin chain, denoted A1. SEQ ID NO: 5 comprises 474 nucleotides.

A particularly advantageous preparation method according to the invention is a method for preparing nucleotide sequences as defined above, characterized in that the pair of primers used is the following: (TGY TGY AGY ATH GAR GAY CG; CA NGC NYC RCT RTT RAA RCA) or (TGY TGY AGY ATH GAR GAY CG; CTC CTC TCC TCT CCT CTT CCT), Y, H, R and N being as defined above,

said method being characterized in that:

– the first stage of the method is a stage of denaturation of the cDNA encoding each of the protein chains constituting the haemoglobin molecule of *Arenicola marina*, of 4 minutes at a temperature equal to 95°C,

– the cycle, repeated 35 times, comprises the following stages:

* a stage of denaturation of the cDNA encoding one of the protein chains constituting the haemoglobin molecule of *Arenicola marina*, of 30 seconds at a temperature equal to 95°C,

* a stage of hybridization of the primer pairs of the invention to the abovementioned strands of monocatenary cDNA in order to obtain hybridized primers, of 40 seconds at a temperature equal to 52°C,

* a stage of elongation of the hybridized primers as obtained previously by a polymerase of 30 seconds at a temperature equal to 72°C, and

– the last stage of the method is a stage of elongation of the hybridized primers as obtained previously by a polymerase of 10 minutes at a temperature equal to 72°C,

in order to obtain the nucleotide sequence SEQ ID NO: 19,

said method optionally comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 7.

The partial sequence SEQ ID NO: 19 was then completed by 5' RACE PCR experiments as explained above. The nucleotide sequence SEQ ID NO: 19 is a novel nucleotide sequence encoding a protein chain corresponding to a globin chain, denoted B2. SEQ ID NO: 19 comprises 390 nucleotides.

The nucleotide sequence SEQ ID NO: 7 (from the start codon to the stop codon, i.e. the transcribed and translated sequence which corresponds to a functional globin monomer) is a novel nucleotide sequence encoding a protein chain corresponding to the abovementioned globin chain, denoted B2. SEQ ID NO: 7 comprises 498 nucleotides.

A particularly advantageous preparation method according to the invention is a method for preparing nucleotide sequences as defined above, characterized in that the pair of primers used is the following: (AAR GTN ATH TTY GGN AGR GA; CTC CTC TCC TCT CCT CTT CCT), R, H, N and Y being as defined above,

said method being characterized in that:

– the first stage of the method is a stage of denaturation of 4 minutes at a temperature equal to 95°C,

– the cycle, repeated 35 times, comprises the following stages:

* a stage of denaturation of 30 seconds at a temperature equal to 95°C,
* a stage of hybridization of 40 seconds at a temperature equal to 52°C,
* a stage of elongation of 30 seconds at a temperature equal to 72°C, and

– the last stage of the method is a stage of elongation of 10 minutes at a temperature equal to 72°C,

in order to obtain a reference partial nucleotide sequence in order to continue the complete determination of this coding sequence,

said method comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 9.

The nucleotide sequence SEQ ID NO: 9 (from the start codon to the stop codon, i.e. the transcribed and translated sequence which corresponds to a functional globin monomer) is a novel nucleotide sequence encoding a protein chain corresponding to a globin chain, denoted B1. SEQ ID NO: 9 comprises 498 nucleotides.

A particularly advantageous preparation method according to the invention is a method for preparing nucleotide sequences as defined above, characterized in that the pair of primers used is the following: (GAR CAY CAR TGY GGN GGN GA, CTC CTC TCC TCT CCT CTT CCT), R, N and Y being as defined above,

said method being characterized in that:

– the first stage of the method is a stage of denaturation of 4 minutes at a temperature equal to 95°C,

– the cycle, repeated 35 times, comprises the following stages:

* a stage of denaturation of 40 seconds at a temperature equal to 95°C,
* a stage of hybridization of 1 minute at a temperature equal to 58°C,
* a stage of elongation of 1 minute and 10 seconds at a temperature equal to 72°C, and

– the last stage of the method is a stage of elongation of 10 minutes at a temperature equal to 72°C,

in order to obtain a reference partial nucleotide sequence in order to continue the complete determination of this coding sequence,

said method comprising an additional stage of 5' RACE PCR in order to obtain the nucleotide sequence SEQ ID NO: 11.

The nucleotide sequence SEQ ID NO: 11 (from the start codon to the stop codon, i.e. the transcribed and translated sequence which corresponds to a functional globin monomer) is a novel nucleotide sequence encoding a protein chain corresponding to a linker chain, denoted L1. SEQ ID NO: 11 comprises 771 nucleotides.

The present invention also relates to protein sequences encoded by one of the nucleotide sequences as obtained according to the method as defined above.

A preferred protein according to the invention is a protein as defined above, characterized in that it comprises or is constituted by:

– the sequence SEQ ID NO: 2 or SEQ ID NO: 14,

– or any sequence derived from the sequence SEQ ID NO: 2 or SEQ ID NO: 14 or from a fragment defined below, in particular by substitution, suppression or addition

of one or more amino acids, provided that said derived sequence allows the transport of oxygen,

– or any sequence homologous to the sequence SEQ ID NO: 2 or SEQ ID NO: 14 or to a fragment defined below, preferably having a homology of at least approximately 75%, in particular of at least approximately 85%, with the sequence SEQ ID NO: 2 or SEQ ID NO: 14, provided that said homologous sequence allows the transport of oxygen,

– or any fragment of one of the sequences defined above, provided that said fragment allows the transport of oxygen, in particular any fragment being made up of at least approximately 60 amino acids, and in particular at least approximately 160 contiguous amino acids in the sequence SEQ ID NO: 2.

The sequence SEQ ID NO: 2 is a novel protein sequence corresponding to a whole globin chain, denoted A2a.

The sequence SEQ ID NO: 14 is a novel protein sequence corresponding to a fragment of a sequence derived from the globin chain, denoted A2a, represented by the sequence SEQ ID NO: 2.

The oxygen transport properties of the protein sequences of the invention can be in particular verified by measuring their absorption spectrum by typical oxyhaemoglobin spectrophotometry.

A preferred protein according to the invention is a protein as defined above, characterized in that it comprises or is constituted by:

– the sequence SEQ ID NO: 4 or SEQ ID NO: 16,

– or any sequence derived from the sequence SEQ ID NO: 4 or SEQ ID NO: 16, or from a fragment defined below, in particular by substitution, suppression or addition of one or more amino acids, provided that said derived sequence allows the transport of oxygen,

– or any sequence homologous to the sequence SEQ ID NO: 4 or SEQ ID NO: 16, or to a fragment defined below, preferably having a homology of at least approximately 75%, in particular of at least approximately 85%, with the sequence SEQ ID NO: 4 or SEQ ID NO: 16, provided that said homologous sequence allows the transport of oxygen,

– or any fragment of one of the sequences defined above, provided that said fragment allows the transport of oxygen, in particular any fragment being made up of at

least approximately 60 amino acids, and in particular of at least approximately 160 contiguous amino acids in the sequence SEQ ID NO: 4.

The sequence SEQ ID NO: 4 is a novel protein sequence corresponding to a whole globin chain, denoted A2b.

The sequence SEQ ID NO: 16 is a novel protein sequence corresponding to a fragment of a sequence derived from the globin chain, denoted A2b, represented by the sequence SEQ ID NO: 4.

A preferred protein according to the invention is a protein as defined above, characterized in that it comprises or is constituted by:

– the sequence SEQ ID NO: 6 or SEQ ID NO: 18,

– or any sequence derived from the sequence SEQ ID NO: 6 or SEQ ID NO: 18 or from a fragment defined below, in particular by substitution, suppression or addition of one or more amino acids, provided that said derived sequence allows the transport of oxygen,

– or any sequence homologous to the sequence SEQ ID NO: 6 or SEQ ID NO: 18 or to a fragment defined below, preferably having a homology of at least approximately 75%, in particular of at least approximately 85%, with the sequence SEQ ID NO: 6 or SEQ ID NO: 18, provided that said homologous sequence allows the transport of oxygen,

– or any fragment of one of the sequences defined above, provided that said fragment allows the transport of oxygen, in particular any fragment being made up of at least approximately 60 amino acids, and in particular of at least approximately 160 contiguous amino acids in the sequence SEQ ID NO: 6.

The sequence SEQ ID NO: 6 is a novel protein sequence corresponding to an entire globin chain, denoted A1.

The sequence SEQ ID NO: 18 is a novel protein sequence corresponding to a fragment of a sequence derived from the globin chain, denoted A1, represented by the sequence SEQ ID NO: 6.

A preferred protein according to the invention is a protein as defined above, characterized in that it comprises or is constituted by:

– the sequence SEQ ID NO: 8 or SEQ ID NO: 20,

– or any sequence derived from the sequence SEQ ID NO: 8 or SEQ ID NO: 20 or from a fragment defined below, in particular by substitution, suppression or addition

of one or more amino acids, provided that said derived sequence allows the transport of oxygen,

– or any sequence homologous to the sequence SEQ ID NO: 8 or SEQ ID NO: 20 or to a fragment defined below, preferably having a homology of at least approximately 75%, in particular of at least approximately 85%, with the sequence SEQ ID NO: 8 or SEQ ID NO: 20, provided that said homologous sequence allows the transport of oxygen,

– or any fragment of one of the sequences defined above, provided that said fragment allows the transport of oxygen, in particular any fragment being made up of at least approximately 60 amino acids, and in particular of at least approximately 160 contiguous amino acids in the sequence SEQ ID NO: 8.

The sequence SEQ ID NO: 8 is a novel protein sequence corresponding to a whole globin chain, denoted B2.

The sequence SEQ ID NO: 20 is a novel protein sequence corresponding to a fragment of a sequence derived from the globin chain, denoted B2, represented by the sequence SEQ ID NO: 8.

A preferred protein according to the invention is a protein as defined above, characterized in that it comprises or is constituted by:

– the sequence SEQ ID NO: 10,

– or any sequence derived from the sequence SEQ ID NO: 10 or from a fragment defined below, in particular by substitution, suppression or addition of one or more amino acids, provided that said derived sequence allows the transport of oxygen,

– or any sequence homologous to the sequence SEQ ID NO: 10 or to a fragment defined below, preferably having a homology of at least approximately 75%, with the sequence SEQ ID NO: 10, provided that said homologous sequence allows the transport of oxygen,

– or any fragment of one of the sequences defined above, provided that said fragment allows the transport of oxygen, in particular any fragment being made up of at least approximately 60 amino acids, and in particular of at least approximately 160 contiguous amino acids in the sequence SEQ ID NO: 10.

The sequence SEQ ID NO: 10 is a novel protein sequence corresponding to a globin chain, denoted B1.

A preferred protein according to the invention is a protein as defined above, characterized in that it comprises or is constituted by:

– the sequence SEQ ID NO: 12,

– or any sequence derived from the sequence SEQ ID NO: 12 or from a fragment defined below, in particular by substitution, suppression or addition of one or more amino acids, provided that said derived sequence allows the combination of globin chains with each other,

– or any sequence homologous to the sequence SEQ ID NO: 12 or to a fragment defined below, preferably having a homology of at least approximately 75%, with the sequence SEQ ID NO: 12, provided that said homologous sequence allows the combination of globin chains with each other,

– or any fragment of one of the sequences defined above, provided that said fragment allows the combination of globin chains with each other, in particular any fragment being made up of at least approximately 60 amino acids, and in particular of at least approximately 280 contiguous amino acids in the sequence SEQ ID NO: 12.

The sequence SEQ ID NO: 12 is a novel protein sequence corresponding to a linker chain, denoted L1.

The present invention also relates to nucleotide sequences as obtained according to the method as defined above.

The present invention also relates to nucleotide sequences encoding a protein as defined above.

The present invention also relates to a nucleotide sequence as defined above, characterized in that it comprises or is constituted by:

– the nucleotide sequence SEQ ID NO: 1 or SEQ ID NO: 13 encoding SEQ ID NO: 2 or SEQ ID NO: 14 respectively,

– or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 1 or SEQ ID NO: 13, and encoding a protein represented by SEQ ID NO: 2 or SEQ ID NO: 14 respectively,

– or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 1 or SEQ ID NO: 13 encoding a protein derived from SEQ ID NO: 2 or SEQ ID NO: 14 respectively,

– or any nucleotide sequence homologous to SEQ ID NO: 1 or SEQ ID NO: 13, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 1,

– or any fragment of the nucleotide sequence SEQ ID NO: 1 or of the nucleotide sequences defined above, said fragment preferably being made up of at least

approximately 180 nucleotides, and in particular of at least approximately 480 contiguous nucleotides in said sequence,

– or any nucleotide sequence complementary to the abovementioned sequences or fragments,

– or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

The stringency conditions correspond to temperature ranges comprised between 48 and 60°C and $MgCl_2$ concentrations comprised between 1 and 3 mM.

The present invention also relates to a nucleotide sequence as defined above, characterized in that it comprises or is constituted by:

– the nucleotide sequence SEQ ID NO: 3 or SEQ ID NO: 15 encoding SEQ ID NO: 4 or SEQ ID NO: 16 respectively,

– or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 3 or SEQ ID NO: 15, and encoding a protein represented by SEQ ID NO: 4 or SEQ ID NO: 16 respectively,

– or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 3 or SEQ ID NO: 15 encoding a protein derived from SEQ ID NO: 4 or SEQ ID NO: 16 respectively,

– or any nucleotide sequence homologous to SEQ ID NO: 3 or SEQ ID NO: 15, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 3 or SEQ ID NO: 15,

– or any fragment of the nucleotide sequence SEQ ID NO: 3 or SEQ ID NO: 15 or of the nucleotide sequences defined above, said fragment preferably being made up of at least approximately 180 nucleotides, and in particular of at least approximately 480 contiguous nucleotides in said sequence,

– or any nucleotide sequence complementary to the abovementioned sequences or fragments,

– or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

The present invention also relates to a nucleotide sequence as defined above, characterized in that it comprises or is constituted by:

– the nucleotide sequence SEQ ID NO: 5 or SEQ ID NO: 17 encoding SEQ ID NO: 6 or SEQ ID NO: 18 respectively,

– or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 5 or SEQ ID NO: 17, and encoding a protein represented by SEQ ID NO: 6 or SEQ ID NO: 18 respectively,

– or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 5 or SEQ ID NO: 17 encoding a protein derived from SEQ ID NO: 6 or SEQ ID NO: 18 respectively,

– or any nucleotide sequence homologous to SEQ ID NO: 5 or SEQ ID NO: 17, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 5 or SEQ ID NO: 17,

– or any fragment of the nucleotide sequence SEQ ID NO: 5 or SEQ ID NO: 17 or of the nucleotide sequences defined above, said fragment preferably being made up of at least approximately 180 nucleotides, and in particular of at least approximately 480 contiguous nucleotides in said sequence,

– or any nucleotide sequence complementary to the abovementioned sequences or fragments,

– or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

The present invention also relates to a nucleotide sequence as defined above, characterized in that it comprises or is constituted by:

– the nucleotide sequence SEQ ID NO: 7 or SEQ ID NO: 19 encoding SEQ ID NO: 8 or SEQ ID NO: 20 respectively,

– or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 7 or SEQ ID NO: 19, and encoding a protein represented by SEQ ID NO: 8 or SEQ ID NO: 20 respectively,

– or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 7 or SEQ ID NO: 19 encoding a protein derived from SEQ ID NO: 8 or SEQ ID NO: 20 respectively,

– or any nucleotide sequence homologous to SEQ ID NO: 7 or SEQ ID NO: 19, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 7,

– or any fragment of the nucleotide sequence SEQ ID NO: 7 or SEQ ID NO: 19 or of the nucleotide sequences defined above, said fragment preferably being made up of at least approximately 180 nucleotides, and in particular of at least approximately 480 contiguous nucleotides in said sequence,

– or any nucleotide sequence complementary to the abovementioned sequences or fragments,

– or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

The present invention also relates to a nucleotide sequence as defined above, characterized in that it comprises or is constituted by:

– the nucleotide sequence SEQ ID NO: 9 encoding SEQ ID NO: 10,

– or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 9, and encoding a protein represented by SEQ ID NO: 10,

– or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 9 encoding a protein derived from SEQ ID NO: 10,

– or any nucleotide sequence homologous to SEQ ID NO: 9, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 9,

– or any fragment of the nucleotide sequence SEQ ID NO: 9 or of the nucleotide sequences defined above, said fragment preferably being made up of at least approximately 180 nucleotides, and in particular of at least approximately 480 contiguous nucleotides in said sequence,

– or any nucleotide sequence complementary to the abovementioned sequences or fragments,

– or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

The present invention also relates to a nucleotide sequence as defined above, characterized in that it comprises or is constituted by:

– the nucleotide sequence SEQ ID NO: 11 encoding SEQ ID NO: 12,

– or any nucleotide sequence derived, by degeneration of the genetic code, from the sequence SEQ ID NO: 11, and encoding a protein represented by SEQ ID NO: 12,

– or any nucleotide sequence derived, in particular by substitution, suppression or addition of one or more nucleotides, from the sequence SEQ ID NO: 11 encoding a protein derived from SEQ ID NO: 12,

– or any nucleotide sequence homologous to SEQ ID NO: 11, preferably having a homology of at least approximately 60%, with the sequence SEQ ID NO: 11,

– or any fragment of the nucleotide sequence SEQ ID NO: 11 or of the nucleotide sequences defined above, said fragment preferably being made up of at least approximately 180 nucleotides, and in particular of at least approximately 800 contiguous nucleotides in said sequence,

– or any nucleotide sequence complementary to the abovementioned sequences or fragments,

– or any nucleotide sequence capable of hybridizing under stringent conditions with the sequence complementary to one of the abovementioned sequences or fragments.

The present invention relates to a preparation method as defined above, for nucleotide sequences encoding the protein chains constituting the haemoglobin molecule of Annelida, in particular of *Arenicola marina*, said method being characterized in that it comprises the following stages:

– a stage of bringing together the abovementioned haemoglobin molecule with at least one dissociating agent and a reducing agent, in particular a mixture made up of dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine hydrochloride (TCEP) or beta-mercaptoethanol and a dissociation buffer, for a sufficient time to separate the protein chains from each other,

allowing the dissociation, then the reduction of said haemoglobin molecule, in order to obtain the protein chains constituting said molecule,

– the isolation of the abovementioned protein chains,

– the microsequencing by mass spectrometry and Edman sequencing of each of the abovementioned isolated protein chains, in order to obtain a microsequence corresponding to each of the sequences made up of 5 to 20 amino acids,

– the determination of the degenerated primer pairs (sense and antisense) from the abovementioned microsequences,

– the preparation of the nucleotide sequences encoding the abovementioned protein chains, from the primers as obtained previously, by a polymerase chain amplification method (PCR), comprising the following stages:

– the first stage of said method is a stage of denaturation of the cDNA encoding the protein chains constituting the haemoglobin molecule of *Arenicola marina*, of approximately 10 seconds to approximately 5 minutes at a temperature comprised between approximately 90°C and approximately 110°C,

– the cycle, repeated approximately 30 to 40 times, comprises the following stages:

* a stage of denaturation of the cDNA encoding the protein chains constituting the haemoglobin molecule of *Arenicola marina*, of approximately 10 seconds to approximately 5 minutes, at a temperature comprised between approximately 90°C and approximately 110°C,
* a stage of hybridization of the primer pairs of the invention to the abovementioned strands of monocatenary cDNA in order to obtain hybridized primers, of approximately 20 seconds to approximately 2 minutes, at a temperature comprised between approximately 50°C and approximately 56°C,
* a stage of elongation of the hybridized primers as obtained previously by a polymerase of approximately 20 seconds to approximately 1 minute and 30 seconds, at a temperature comprised between approximately 70°C and approximately 75°C, and

– the last stage of the method is a stage of elongation of the hybridized primers as obtained previously by a polymerase of approximately 5 minutes to approximately 15 minutes at a temperature comprised between approximately 70°C and approximately 75°C.

DESCRIPTION OF THE FIGURES

Figure 1 represents a chromatogram of the haemoglobin of *Arenicola marina* on a Superose 12-C column. The upper curve corresponds to an absorbance of 414 nm and the lower curve to an absorbance of 280 nm. (The collector is programmed to collect between 16 and 18 minutes).

Figure 2 represents the UV spectrum of the functional haemoglobin of *Arenicola marina* (in its oxyhaemoglobin form).

Figure 3 represents the chromatogram (at 414 nm) of the (partially) dissociated HbAm obtained on Superose 12-C and the vertical lines on the chromatogram correspond to the collecting windows (corresponding to the recovery of the sub-units).

Figure 4 represents an SDS-PAGE gel obtained for the different fractions collected.

Figure 5 represents the chromatogram (at 414 nm) of the (partially) dissociated HbAm obtained on CIM DISK DEAE (anionic exchange system) and the vertical lines on the chromatogram correspond to the collecting windows. The dotted curve indicates the gradient.

Figure 6 represents an SDS-PAGE gel obtained for the different fractions collected.

Figure 7 represents the dissociation kinetics of the HbAm in the presence of 3M urea. The x-axis corresponds to the number of days and the y-axis corresponds to the percentage of dissociation of the native molecule; the dotted curve corresponds to the dodecamer; the curve with the black squares to the trimer and the "linker" (structural chain); the curve with the black circles to the monomers.

Figure 8 represents the dissociation kinetics of the HbAm at pH 10. The x-axis corresponds to the number of days and the y-axis corresponds to the percentage of dissociation of the native molecule; the dotted curve corresponds to the dodecamer; the curve with the black squares to the trimer and the "linker"; the curve with the black circles to the monomers.

Figure 9 represents the dissociation kinetics of the HbAm in the presence of 3M urea at pH 10. The x-axis corresponds to the number of days and the y-axis corresponds to the percentage of dissociation of the native molecule; the dotted curve corresponds to the dodecamer; the curve with the black squares to the trimer and the "linker"; the curve with the black circles to the monomers.

Figure 10 represents the monitoring of the reassociation kinetics from the percentage of HbAm (HBL) and Dodecamer (D) and according to the buffer changing technique (Centricon or Dialysis). The x-axis corresponds to the number of days and the y-axis corresponds to the percentage of dissociation of the native molecule with the Centricon technique; the dotted curve corresponds to HBL with the dialysis technique; the curve with the black triangles corresponds to the dodecamer with the Centricon technique; the full line curve corresponds to the dodecamer with the dialysis technique.

Figure 11 represents the superposition of the exclusion chromatography chromatograms during the reassociation corresponding to different reassociation times.

Figure 12 represents the HPLC chromatogram obtained after separation of the polypeptide chains by Reversed-phase on a Symmetry C18 column (Waters). The codes (d2, a1, a2, b2, c) correspond to the names of the globins as mentioned in the article by Zal et al. (1997).

EXPERIMENTAL PART

An objective of the present invention is the use of the extracellular haemoglobin of the marine polychaete *Arenicola marina* (HbAm) as a blood substitute in vertebrates. However, even if this worm represents a significant biomass, synthesis by genetic engineering has proved to be an indispensable and necessary route. It is therefore of prime importance to obtain the primary sequences of the protein chains constituting HbAm in order to develop an artificial, functional and stable haemoglobin from the self-assembly properties of this molecule. The dissociation protocols of each sub-unit and reduction to polypeptide chains, as well as the purification, isolation, microsequencing and sequencing techniques of these chains are discussed in detail hereafter.

Extraction and purification of the haemoglobin

1) Species studied: *Arenicola marina*; Annelida of the intertidal ecosystem

The Annelida Polychaete *Arenicola marina* is a sedentary species widespread throughout all the coasts of the North Atlantic, Black Sea and Adriatic situated above the fortieth parallel. In the Roscoff region, the Arenicola, commonly known in French as the "ver du pêcheur" forms dense populations. The sediment inhabited by these populations has an irregular surface of alternating bumps and hollows formed respectively by mounds of coprogenous particles and conical depressions. The Arenicola lives in galleries made in the sand. The structure of the gallery is presented in the shape of a U, with an open branch on the outside, the other being closed. The Arenicola is accommodated in the horizontal part, its cephalic end oriented towards the blind part. It ingests the sand, extracts the assimilable organic matter and then defecates through its caudal end, thus forming of the mounds of wormcasts of sand. Inside the mediolittoral stage, the distribution and density of the populations are essentially controlled by the granulometry, the concentration of organic matter and the salinity.

The Arenicola, living above all in the tidal zones, has to undergo variations in oxygen pressure. Its gallery makes it possible for it to be in permanent contact with seawater (rich in oxygen) at low tide.

2) Methods of study

2.1. Sampling of the biological material

The animals are collected at low tide, in the Baie de Penpoull, near Roscoff (France) and kept in seawater overnight in order to empty their digestive tube. The blood samples are taken from the dorsal vessel using a syringe. The samples are collected on ice and filtered through glass wool. After low-temperature centrifugation (15,000 g for 15 min at 4°C) in order to avoid the dissociation of the molecules and eliminate the tissue debris, the supernatants are concentrated by means of an Amicon cell (Millipore) and a "cut-off" membrane of 500 kDa (only masses greater than 500 kDa are retained).

2.2. Purification of the haemoglobins

Once the blood is concentrated, a low-pressure filtration (FPLC) by exclusion (separation as a function of the size of the molecule: the more significant the size of the molecule the more rapidly it is eluted) is carried out on a column (100 × 3 cm) of Sephacryl S-400 gel (Amersham)(separation range comprised between 20×10^3 and 8000×10^3), in a cold room (4°C). Each purification is carried out on 5 mL of sample, eluted with the *Arenicola marina* salinated buffer (10 mM Hepes; 4 mM KCl; 145 mM NaCl; 0.2 mM $MgCl_2$ adjusted to pH 7.0 with 2N soda). The flow rate used for this first purification is 40 r.p.m. and only the first, reddest, fraction (containing heme) is recovered. This fraction is then concentrated using a Centricon-10 kDa tube retaining the molecules with a weight greater than or equal to 10,000 Da.

A second purification is then carried out by low-pressure filtration (HPLC System, Waters) of 200 µL aliquots on a 1 × 30 cm Superose 12-C column (Pharmacia, separation range comprised between 5×10^3 and 3×10^5 Da) at ambient temperature. The flow rate used is 0.5 ml/min. The samples are kept at 4°C and collected in ice. The absorbance of the eluate is monitored at two wavelengths: 280 nm (absorbance peak characteristic of proteins) and 414 nm (absorbance peak characteristic of heme). The fractions containing heme are isolated using a collector (programmed on a time window corresponding to the retention time of the haemoglobin) (Figure 1). The samples are concentrated, assayed, then stored at -40°C before use.

2.3. Assay of the haemoglobins

Drabkin's reagent (Sigma), used for the assay, makes it possible to determine the quantity of heme in the solution. The haemoglobin reacts with Drabkin's reagent which contains potassium ferricyanide, potassium cyanide and sodium bicarbonate. The haemoglobin is converted to methaemoglobin by the action of the ferricyanide. The methaemoglobins then react with the cyanide in order to form cyanmethaemoglobin. The absorbance of this derivative at 540 nm is proportional to the quantity of heme in the solution. The extracellular haemoglobin of *Arenicola marina* (HBL) contains on average 1 mol of heme per 23,000 g of protein, which makes it possible by a simple calculation, to obtain the HBL concentration of each sample.

3) Results

Thus, several milligrams of extracellular haemoglobin of *Arenicola marina* were purified. Each batch (1 mL aliquots) is analyzed by FPLC on a Superose 12-C column (Pharmacia) in order to ensure the purity of the sample (a single peak). Similarly a UV spectrum over a range of 400 nm to 700 nm is produced in order to verify the functionality of the haemoglobins of each batch (Figure 2). Three absorption maxima are observed at 414, 541 and 577 nm. By comparison, it should be recalled that the methaemoglobin exhibits two maxima at 500 and 635 nm.

Finally, these batches are used by Biotrial S.A. (Rennes, France) for preclinical tests carried out on mice and rats in order to test the possible pathological and immunogenic reactions.

Dissociation of the extracellular haemoglobin of *Arenicola marina* in these different basic sub-units (trimers, linker dimers and monomers)

The dissociation of the extracellular haemoglobin of *Arenicola marina* (HbAm) must be total and retain the functional sub-units. The different sub-units are then isolated and analyzed by the liquid chromatography technique (exclusion and ion exchange), developed for this purpose.

1) Dissociation of the HBL

1.1. Dissociation protocol

The preliminary studies of dissociation were developed from the publications of the prior art (Vinogradov et al., 1979; Sharma et al., 1996; Mainwaring et al., 1986; Polidori et al., 1984; Kapp et al., 1984; Chiancone et al., 1972; Vinogradov et al., 1991; Krebs et al., 1996), i.e. in the presence of a single dissociating reagent: urea, heteropolytungstate ions, guanidinium salts, SDS or hydroxide ions. The agents used act differently on the molecule:

– the hydroxide ions (OH^-), the SDS, the guanidinium salts and the heteropolytungstate ions destabilize the salt bridges

– the urea destabilizes the hydrophobic interactions

The aim is to obtain the four basic sub-units as rapidly and effectively as possible, hence the idea of combining the different dissociating agents and in particular the alkaline pH and urea. After different tests, it emerges that a rapid and effective dissociation is obtained with 3M urea diluted in the dissociation buffer (0.1 M of Trisma base and 1 mM of EDTA) adjusted to pH 10 with 2N soda. The HbAm is adjusted to a concentration of approximately 4 mg/mL (stock solution). All the analyses are carried out at +4°C and the samples are kept in the dark throughout the study. (Trisma = tris[hydroxymethyl]aminomethane)

1.2. Exclusion chromatography analyses

The analysis conditions are shown in detail in Table 1 below.

System HPLC	HPLC Waters 626 LC System	
Column	Superose 12-C (Pharmacia) (separation range comprised between $5x10^3$ and $3x10^5$ Da)	
Flow rate	0.5 mL/min	
Eluent	buffer pH 7.0 *(buffered for example with concentrated hydrochloric acid)* and filtered through a 0.22 µm (or 0.45µm) filter	
Temperature of the injecter	+ 4°C	
	Sample Analytic monitoring	Separation and collection
Volume injected	20µL	200µL
Preparation of the samples	Stock solution diluted to 1mg/mL in the dissociation buffer at pH 10 and filtered through 0.45µm	Stock solution filtered through a 0.45µm filter

1.3. Ion-exchange analyses

The isoelectric point (pHi) of HBL being 4.69 (Vinogradov, 1985), ion-exchange analysis is carried out on a CIM DEAE disk anionic column (Interchim). In fact, HBL is negatively charged for a pH greater than the pHi and is therefore fixed on a positively charged resin (DEAE resin). The elution is carried out by means of the ionic force with a non-linear NaCl gradient of 0 to 1 M (1 M NaCl solution diluted in the dissociation buffer at pH 7.0 and filtered through 0.45 µm). The dissociation buffer at pH 7 is used as elution buffer. The flow rate is 4 mL/min.

1.4. Reversed-phase chromatography analyses

Reversed-phase chromatography is carried out on a Waters 300 C_{18} 5µm (4.6 × 250 mm) Symmetry column. In the presence of acetonitrile and TFA (trifluoroacetic acid), HbAm is dissociated into its basic sub-units (Trimer, Monomer and Linker) and the heme is dissociated from the globins. Thus, without previous treatment, HbAm is dissociated at the column head. The method developed is described in the table below.

flow rate	1 mL/min		
Eluent A	H_2O MilliQ + 0.1% v/v HFBA		
Eluent B	ACN + 0.1% v/v HFBA		
Gradient			
	Time in min	% A	% B
	0	58	42
	40	50	50
	41	48	52
	90	47	53
	95	5	95
	110	5	95

2) Results

2.1. Exclusion chromatography

The chromatogram of the partly dissociated HbAm is represented in Figure 3. Five peaks are observed and have to be identified. The molecules are eluted according to their decreasing mass and the eluted native HbAm in the hold-up volume (16 min). The fractions corresponding to each peak are analyzed by SDS-PAGE (Figure 4). The results are presented in Table 2 below.

Fractions	Retention time	Sub-units
1	16 min 40	Native HbAm
2	22 min 20	Dodecamer
3	25 min 30	Linker dimer
4	26 min 40	Trimer
5	28 min 30	Monomers

2.2. Ion chromatography

Once the method has been developed (Table 3), the fractions are collected, concentrated and analyzed by SDS gel in order to identify each peak (Fig. 6 and Table 4).

A method is then developed for repurifying each sub-unit.

Time in min	A : 1M NaCl in B	B: Dissociation buffer pH 7.0
0	5%	95 %
0.5	15%	85 %
1.5	15%	85 %
2.5	22%	78 %
3.5	22%	78 %
3.6	25%	75 %
5.5	25%	75 %
5.6	29%	71 %
6.5	29%	71 %
6.6	36%	64 %
7.0	36%	64 %
8.0	45%	55 %
8.1	100%	0 %
9.5	100%	0 %

TABLE 3: Method developed for the analysis of the dissociated HBL on CIM-DEAE

fractions	Retention time	Sub-units
1	1 min 15	Monomers
2	2 min 40	Linker dimer
3	3 min 10	Linker dimer
4	4 min 40	Dodecamer
5	6 min 30	?
6	7 min 30	Trimer
7	8 min 50	HbAm

TABLE 4: Association carried out after analysis of the gel (Figure 6) between the retention time and the sub-units.

2.3. Reversed-phase chromatography

Once the method has been developed, the fractions are collected, lyophilized and analyzed by mass spectrometry in order to identify each peak.

fractions	Retention time	Sub-units
1	12 min	Linker
2	22 min	Linker
3	34 min	Monomer a1
4	38 min	Monomer a2
5	50 -70 min	Trimers

The chemical properties of the trimers must be too close for it to be possible to separate them by reversed-phase. Thus, it has been possible to isolate only the two linkers and the monomers a1 and a2.

2.4. Dissociation of the HbAm

The dissociation kinetics are monitored by exclusion chromatography. The integration of the chromatograms by Millenium software (Waters) makes it possible to calculate the percentage of the different compounds from the area under the curve. The evolution of the dissociation kinetics is represented in Figures 7, 8 and 9.

The three graphs in Figures 7, 8 and 9 show the benefit of combining the two dissociating agents (3 M urea and OH^-) in order to effectively obtain the three basic sub-units in 24 hours.

Reassociation of the haemoglobin

1) Materials and methods

The reassociation experiments are carried out on dissociated HbAm according to the protocols mentioned above (pH9, pH10, 3M Urea, 4M Urea, 3M Urea at pH 10). Different reassociation buffers are tested in order to obtain an optimum reassociation. The change of buffer (dissociation buffer → reassociation buffer) is carried out in two different ways:

– The dissociated HbAm is washed 4 times against 4 mL of reassociation buffer on Centricon-10 (Millipore) at +4°C;

– The dissociated HbAm is dialyzed for 24 hours against MilliQ water (Millipore) (2 × 2L) then for 48 hours against the reassociation buffer (3 × 2L) at +4°C.

2) Results

According to subsequent results relating to the extracellular haemoglobins of Annelida (Mainwaring et al., 1986; Polidori et al., 1984), the presence of divalent ions such as Ca^{2+} and Mg^{2+} is necessary for maintenance of the quaternary structure of haemoglobin. In fact, they stabilize it and slow down the dissociation phenomenon (Sharma et al., 1996). These ions form a complex with the carboxylate groups of the side chains and carbonyls of the main chains. The presence of divalent ions can have an effect on the reassociation when the carboxylic groups are ionized, therefore *inter alia* when the dissociation has taken place at an alkaline pH. It is therefore significant that the reassociation buffer contains calcium and/or magnesium. This also explains the presence of EDTA in the dissociation buffer; EDTA which chelates these divalent ions. The buffer now developed is made up of 0.1 M of Trisma base, 400 mM of NaCl, 2.95 mM KCl, 32 mM $MgSO_4$, 11 mM $CaCl_2$ adjusted to pH 7 with concentrated HCl. The reassociation is monitored according to the same principle as the dissociation (Figures 10 and 11).

A reassociation is observed if the dissociation is of short duration of the order of one minute. This reassociation corresponds to a rearrangement of dissociation intermediates which are truncated haemoglobins (HBL dissociated from 1 or more twelfths).

Reduction of HbAm for the study of the different polypeptide chains

1) Reduction of haemoglobin prior to separation by reversed-phase liquid chromatography

The HbAm (4 mg/mL) is reduced in 10% DTT (dithiothreitol) dissolved in a dissociation buffer at pH 8-9 (0.1 M trisma or 10 mM ammonium bicarbonate) for 30 minutes at ambient temperature. Once reduced, the protein chains obtained are alkylated in the presence of 100 mM iodoacetamide dissolved in a dissociation buffer at pH 8-9 for 30 minutes at ambient temperature.

The following protocol can also optionally be envisaged: The HbAm (4 mg/mL) is reduced in 100 mM of DTT (dithiothreitol) dissolved in the dissociation buffer at pH 8-9 for 1 hour at 40°C. Under these drastic conditions, only the globins can be analyzed.

In fact, the linkers (non-globin proteins) are rich in cysteines and are therefore damaged. Once reduced, the HbAm is washed 4 times on Amicon 30,000 Da (Millipore) only the filtrate of which is recovered (all which has a weight of less than 30,000 Da). The filtrate is then washed on Amicon 10,000 Da in order to eliminate all which is less than 10,000 Da. Thus, only the monomers comprised between 30,000 Da and 10,000 Da are contained in the sample (weight range of the globin chains which constitute HbAm).

2) Separation of the protein chains by reversed-phase chromatography

2.1. Materials and methods

Reversed-phase chromatography is carried out on a Waters Symmetry 300 C_{18} 5µm (4.6 × 250 mm) column. The method developed is described in the table below and the chromatogram obtained is represented in Figure 12.

flow rate	1 mL/min		
Eluent A	H_2O MilliQ + 0.1% v/v HFBA		
Eluent B	ACN + 0.1% v/v HFBA		
	Gradient		
	Time in min	% A	% B
	0	75	25
	2	58	42
	10	58	42
	40	40	60
	45	0	100
	55	0	100

The following protocol can also be optionally envisaged.

flow rate	1 mL/min	
Eluent A	H_2O MilliQ + 0.1% v/v TFA	
Eluent B	ACN + 0.1% v/v TFA	
Gradient		
Time in min	% A	% B
0	80	20
0.25	60	40
4.0	55	45
10.0	55	45
10.05	0	100
16.0	0	100

Each protein chain (revealed by a single peak at 280 nm) is collected then lyophilized and stored at -40°C until the next analyses. Thus, it has been possible to separate the following 5 monomers: a_1 (~15952 kDa), a_2 (~15975 kDa), d_2 (~17033 kDa), b_2 (~16020 kDa) and c (~16664 kDa).

3) Separation of the protein chains by two-dimensional gel

Two-dimensional gel which is a combination of the isoelectric focusing technique in the first dimension and SDS-PAGE technique in the second dimension makes it possible to separate a complex protein mixture.

Isoelectric focusing

After purification by FPLC, the haemoglobin of *Arenicola* is dialyzed and lyophilized. 500 µg are taken up in a rehydration buffer. This buffer contains 4% Chaps (3-[(3-cholamidopropyl)dimethylammonio]-1-propane-sulphonate; Sigma), 1% freshly prepared DTT, a cocktail of protease inhibitors (Bohringher), 50 µg/ml of TLCK (trypsin inhibitor) and 1 µl of 1% Bromophenol Blue.

The mixture is sonicated and centrifuged in order to eliminate the non-dissolved material. Stone oil is then applied to the two ends of the support of the isoelectric focusing band, and the sample is then applied to the medium. The 17 cm band is then applied to the sample, eliminating any air bubbles. The band is then covered with stone oil in order to avoid evaporation of the sample.

An active rehydration is then carried out at 50V (20°C over 12 hours). The focalization is then carried out over two days.

The band is then recovered and placed on a 6-18% acrylamide gel, in particular 10%, 18 cm wide, 20 cm long and 1 mm thick. The migration is carried out in a refrigerated enclosure at 10°C, over 14 hours at 400 V, 25 mA and 100 W.

The separation of the protein chains is then carried out as a function of their size after sealing the band on top of the gel using a 1% agarose solution. Once separated, the protein bands are revealed on gel by staining with Coomassie blue (Coomassie® G250).

Construction of the gradient gel

Twenty-five ml of 18% polyacrylamide dense solution (2.5 M acrylamide; 0.4 M Tris; 30% Glycerol (v/v); 3.5 mM sodium dodecyl sulphate (SDS); 0.05% TEMED (N,N,N',N'-tetramethylethylenediamine; Sigma) (v/v); 1.6 mM sodium persulphate) are placed in a mixing chamber under constant stirring whilst the same volume of 6% polyacrylamide light solution (acrylamide 0.8 M; Tris 0.4 M; SDS 3.5 mM; TEMED 0.06% (v/v); sodium persulphate 2.4 mM) is placed in the other chamber. The top of the gel is covered with a saturated isobutanol solution in bidistilled water. The gel is then left for 1 hour to polymerize at ambient temperature, then the top of the gel is rinsed

several times with bidistilled water and the whole is placed overnight at 10°C. After removal of the residual water using an absorbent paper, the concentration gel solution (0.56 M acrylamide; 6.9 mM methylene bis-acrylamide; 124 mM Tris; 3.5 mM SDS; 0.05% TEMED (v/v); 2.2 mM sodium persulphate) is poured onto the separation gel and a shim making it possible to form the blot of the band is introduced into the concentration gel solution. The polymerization is complete after 1 hour at ambient temperature.

4) Analysis by micro sequencing of the isolated protein chains by reversed-phase and two-dimensional gel

The protein chains isolated by reversed-phase chromatography and by two-dimensional gel are then digested and analyzed by LC-MS/MS mass spectrometry on an ESI-Q-TOF type device.

Digestion of the separated proteins by reversed-phase chromatography.

Each isolated protein chain is subjected to enzymatic digestion, an essential stage before their analysis by microsequencing. The lyophilized protein chains are dissolved in a milliQ water solution, acetonitrile containing endoprotease; trypsin which hydrolyzes at the C-terminal level of lysine and arginine, generally producing peptides with masses comprised between 500 and 2500 Da, over a minimum of 3 hours at ambient temperature.

Digestion of the separated proteins on two-dimensional gel

Each spot of the gel is cut out in order to be subjected to enzymatic digestion. This enzymatic digestion stage is essential. It consists of hydrolyzing the proteins in a specific manner, using an enzyme, into several peptides.

Before beginning digestion, discoloration, reduction and alkylation stages are indispensable:

- successive washings with ammonium hydrogen carbonate (NH_4HCO_3) and acetonitrile (ACN) make it possible to eliminate the staining agent present in the piece of gel,
- the reduction reactions with dithiothreitol (DTT) and alkylation reactions with iodoacetamide allow the opening then the blocking of the disulphide bridges formed between two cysteines present in the protein sequence and cysteine-acrylamide bonds.

The last stage of the method consists of extracting the trypsic peptides from the gel using an extraction solution, composed of acetonitrile and water, with a little acid added.

Analysis by nanoLC-MS-MS

The extracted peptides are then transferred to the PCR plate. This transfer is carried out with twice 15 µL, in order to recover all of the volume. In order to eliminate the acetonitrile, which could impede the retention of the peptides on the pre-column, a time of evaporation (pause) of 2 hours is applied before analysis by nanoLC-MS-MS.

Results

This made it possible to obtain a few hundred micosequences corresponding to each separate chain protein by 2D gel.

5) Analysis by Edman sequencing of the isolated protein chains by reversed-phase chromatography

Principle

In the presence of N-methyl piperidine buffer, Phenyl-Iso-Thio-Cyanate (PITC) is coupled to the primary and secondary amine functions of the proteins (PTC-Protein). The reaction time at 45°C is 18 minutes. The following peptide bond is weakened, which allows it to be cut in 3 minutes by pure trifluoroacetic acid (TFA) thus generating the anilino-thiazolinone (ATZ) of the first amino acid (AA) and the protein having lost the 1st AA.

The ATZ-AA is extracted from the reaction medium and converted in acid medium (25 % TFA in water) to the more stable phenyl thio-hydantoin (PTH-AA). The PTH-AA can therefore be analyzed by HPLC and its nature determined by means of a PTH-AA standard. The reaction cycle can be repeated and thus leads to the protein sequence. Edman automated the reaction which bears his name by creating the first protein sequencer in 1967. The device is coupled to an HPLC into which it injects PTH-AA. By comparison with a standard spectrum, it is then possible to identify the original amino acid and obtain its quantification. The whole process is controlled by a computer

which controls the different elements and ensures the acquisition of data as well as their processing.

Results

Thus, it was possible to obtain approximately 30 amino acids from the N-terminal ends of 5 monomers and an linker isolated by reversed-phase.

Details of PCR amplification protocols and presentation of the nucleotide and polypeptide sequences of the globins of the sub-families A1, A2a, A2b, B2 and B1 and the linker L1 of the marine polychaete *Arenicola marina*

The PCR amplifications of the 5 globins A1, A2a, A2b, B1 and B2, as well as of the linker L1, the nucleotide sequences of which are presented below, commenced with the design of specific degenerated primers (sense and antisense) of the sub-families A1, A2, B1 and B2. These primers, which allowed the amplification of the abovementioned five globins (A1, A2a, A2b, B1 and B2) then the cloning and sequencing of the corresponding PCR products, were designed from alignments of protein sequences of Annelida globins available from the data banks.

The complementary DNA matrices used for the PCR reactions were synthesized from messenger RNAs purified from total RNAs extracted from Arenicolas, due to the small size of the organisms and their intense growth rate reflecting significant levels of expression of the genes, including those involved in the synthesis of the haemoglobin. The complementary DNAs have thus been synthesized. These stages made use of commercial molecular biology kits produced by Ambion (purification of the RNAs), Amersham (purification of the mRNAs), Promega (RT), Invitrogene (cloning), Abgene (sequencing).

In a second phase, we developed the PCR reactions, in particular as regards the determination of the denaturation time, hybridization time and temperature and elongation time parameters. The $MgCl_2$ concentrations were also optimized.

Finally, in a last stage, 5' and 3' RACE PCR experiments were carried out so as to obtain the complete coding sequences. These stages used the Roche molecular biology kit.

The nucleotide sequences of the degenerated sense and antisense primers, the PCR parameters and the partial or complete coding sequences for each of the globins A2a, A2b, A1, B1 and B2, and for the linker L1 are presented below.

It is specified that the total blast databank analysis of these sequences produces values comprised between $2.10^{-3} < \text{Evaluate} < 5^{e-31}$.

Globin A2a

In order to obtain the nucleotide sequence SEQ ID NO: 1 encoding the globin A2a (SEQ ID NO: 2), the pair of primers (SEQ ID NO: 19; SEQ ID NO: 20) are used.

The PCR conditions are the following:

Time and initial temperature of denaturation:	4 min at 95°C	
Time and Temperature of denaturation:	30 s at 95°C	35 cycles
Time and Temperature of hybridization:	30 s at 56°C	
Time and Temperature of elongation:	40 s at 72°C	
Time and Temperature of final elongation:	10 min at 72°C	

PCR Reaction:

Per reaction:	5-20 ng cDNA
	100 ng sense primer
	100 ng antisense primer
	dNTP 200 µM final
	$MgCl_2$ 2 mM final
	Buffer PCR 1X final
	1 unit Taq Polymerase
	Qsf 25 µL H_2O

Globin A2b

In order to obtain the nucleotide sequence SEQ ID NO: 3 encoding the globin A2b (SEQ ID NO: 4), the pair of primers (SEQ ID NO: 21; SEQ ID NO: 20) are used.

The PCR conditions are the following:

PCR:

4 min at 95°C	
30 s at 95°C	35 cycles
30 s at 52°C	
40 s at 72°C	

10 min at 72°C

Globin A1

In order to obtain the nucleotide sequence SEQ ID NO: 5 encoding the globin A1 (SEQ ID NO: 6), the pair of primers (SEQ ID NO: 22; SEQ ID NO: 20) are used.

The PCR conditions are the following:

PCR: 4 min at 95°C

1 min at 95°C
1 min at 50°C
1 min 30 at 72°C
} 35 cycles

10 min at 72°C

Globin B2

In order to obtain the nucleotide sequence SEQ ID NO: 7 encoding the globin B2 (SEQ ID NO: 8), the pair of primers (SEQ ID NO: 23; SEQ ID NO: 20) are used.

The PCR conditions are the following:

PCR: 4 min at 95°C

30 s at 95°C
40 s at 52°C
30 s at 72°C
} 35 cycles

10 min at 72°C

Globin B1

In order to obtain the nucleotide sequence SEQ ID NO: 9 encoding the globin B1 (SEQ ID NO: 10), the pair of primers (SEQ ID NO: 24; SEQ ID NO: 20) are used.

The PCR conditions are the following:

PCR: 4 min at 95°C

30 s at 95°C
40 s at 52°C
30 s at 72°C
} 35 cycles

10 min at 72°C

Linker L1

In order to obtain the nucleotide sequence SEQ ID NO: 11 encoding the Linker L1 (SEQ ID NO: 12), the pair of primers (SEQ ID NO: 25; SEQ ID NO: 20) are used.

The PCR conditions are the following:

PCR: 4 min at 95°C

40 s at 95°C
1 min at 58°C
1 min at 72°C
} 35 cycles

10 min at 72°C

BIBLIOGRAPHICAL REFERENCES

- Bunn, H.F. et Jandl, J.H. (1968) *Trans Assoc Am Physicians*, **81**, 147,
- Chang, T.M.S. (1957) Hemoglobin Corpuscles, McGill University,
- Chang, T.M.S. (1964) *Science*, **146**, 524-525,
- Chang, T.M.S. (1971) *Biochem. Biophys. Res. Com.*, **44**, 1531-1533,
- Chang, T.M.S. (1997) Blood substitutes: principals, methods, products and clinical trials, Vol. I, Karger, Basel, Suisse,
- Chiancone, E., et al. (1972) Studies on erythrocruorin. II. Dissociation of earthworm erythrocruorin, *J Mol Biol*, **70**(1): 73-84,
- Clark, L.C.J. et Gollan, F. (1966) *Science*, **152**, 1755,
- De Haas, F.; Boisset, N.; Taveau, J.C.; Lambert, O.; Vinogradov, S.N. and Lamy, J.N. (1996) *Biophys. J.*, **70**, 1973-1984,
- De Haas, F.; Taveau, J.-C.; Boisset, N.; Lambert, O.; Vinogradov, S.N. and Lamy, J.N. (1996) *J. Mol. Biol.*, **255**, 140-153,
- De Haas, F.; Zal, F.; Lallier, F.H.; Toulmond, A.and Lamy, J.N. (1996) Proteins-structure fonction and genetics, **3**, 241-256,
- De Haas, F.; Zal, F.; You, V.; Lallier, F.H.; Toulmond, A. and Lamy, J.N. (1996) *J. Mol. Biol.*, 264, 111-120,
- Dickerson, R.E. and Geis, I. (1983) Hemoglobin: Structure, function, evolution, and pathology., Benjamin/Cummings, Menlo Park, CA,
- Geyer, R.P.; Monroe, R.G. and Taylor, K. (1968) Survival of rats totally perfused with a fluorocarbon-detergent preparation., J.C. Norman, J. Folkman, W.G. Hardisson, L.E. Rudolf and F.J. Veith (Eds), 85-96, Appleton Century Crofts, New York,
- Goodin, T.H.; Grossbard, E.B.; Kaufman, R.J.; Richard, T.J.; Kolata, R.J.; Allen, J.S. and Layton, T.E. (1994), *Crit Care Med*, **22**, 680-689,
- Hirsch, R.E.; Jelicks, L.A.; Wittenberg, B.A.; Kaul, D.K.; Shear, H.L. and Harrington, J.P. (1997) Artificial Cells, Blood Substitutes & Immobilization Biotechnology, **25**, 429-444,
- Jia, L.; Bonaventura, C.; Bonaventura, J. and Stamler, J.S. (1996) *Nature*, **380**, 221-226,
- Kapp, O.H. and Crewe, A.V. (1984) *Biochim. Biophys. Acta* **789**, 294-301,

– Kapp, O.H., et al. (1984) The reassociation of Lumbricus terrestris hemoglobin dissociated at alkaline pH, *J Biol Chem*, **259**(1): 628-39,

– Krebs, A., et al. (1996) Molecular shape, dissociation, and oxygen binding of the dodecamer subunit of Lumbricus terrestris hemoglobin, *J Biol Chem*, **271**(31): 18695-704,

– Lamy, J.N.; Green, B.N.; Toulmond, A., Wall, J.S.; Weber, R.E. and Vinogradov, S.N. (1996), *Chem. Rev.* **96**, 3113-3124

– Levin, O. (1963) *J. Mol. Biol.*, **6**, 95-101,

– Mainwaring, M.G., et al. (1986) The dissociation of the extracellular hemoglobin of Lumbricus terrestris at acid pH and its reassociation at neutral pH. A new model of its quaternary structure, *J Biol Chem*, **261**(23): 10899-908,

– Mitsuno, T. and Naito, R. (1979) Perfluorochemical Blood Substitutes., Excerpta Medica, Amsterdam,

– Mitsuno, T. and Ohyanagi, H. (1985) Present status of clinical studies of fluosol-DA (20%) in Japan, K.K. Tremper (Ed), 169-184, Little Brown & Co, Boston,

– Naito, R. and Y. K. (1978) An improved perfluorodecalin emulsion. In Blood Substitutes and Plasma Expanders, G.A. Jamieson and T.J. Greenwalt (Eds), 81, Alan R. Liss Inc., New York,

– Nho, K.; Glower, D.; Bredehoeft, S.; Shankar, H.; Shorr, R. and Abuchowski, A. (1992) Biomaterials, Artificial Cells and Immobilization Biotechnology, An International Journal, **20**, 511-524,

– Payne, J.W. (1973) *Biochem J.* **135**, 866-873,

– Polidori, G., et al. (1984) The dissociation of the extracellular hemoglobin of Tubifex tubifex at extremes of pH and its reassociation upon return to neutrality, *Arch Biochem Biophys*, **233**(2): 800-14,

– Reiss, J.G. (1991) *Vox Sang*, **61**, 225-239,

– Reiss, J.G., (1994) Artificial Cells, Blood substitutes & Immobilization Biotechnology, An International Journal **22**, 945-1511,

– Roche, J. (1965) Electron microscope studies on high molecular weight erythrocruorins (invertebrate hemoglobins) and chlorocruorins of annelids., D.A. Munday (Ed), 62-80, Pergamon Press, Oxford,

– Roche, J.; Bessis, M. and Thiery, J.P. (1960) *Biochim. Biophys. Acta*, **41**, 182-184,

– Roche, J.; Bessis, M.T. and Thiery, J.P. (1960) *C. R. Soc. Biol.* **154**, 73-80,

– Sharma, P.K., et al. (1996) The role of the dodecamer subunit in the dissociation and reassembly of the hexagonal bilayer structure of Lumbricus terrestris hemoglobin, *J Biol Chem*, **271**(15): 8754-62,

– Sloviter, H. and Kamimoto, T. (1967) *Nature*, **216**, 458,

– Terwilliger, N.B. (1992) Molecular Structure of the extracellular heme proteins., Vol. 13, C.P. Mangum (Ed), 193-229, Springer-Verlag, Berlin,

– Van Bruggen, E.F.J. and Weber, R.E. (1974) *Biochim. Biophys. Acta*, **359**, 210-212,

– Vinogradov, S.N., et al. (1991) A dodecamer of globin chains is the principal functional subunit of the extracellular hemoglobin of Lumbricus terrestris, *J Biol Chem*, **266**(20): 13091-6,

– Vinogradov, S.N., Shlom J. M. and Doyle, M. (1979) Dissociation of the extracellular hemoglobin of Arenicola cristata. *Comp. Biochem. Physiol.*, **65B**: 145-150,

– Vinogradov, S.N. (1985) The structure of invertebrate extracellular hemoglobins (erythrocruorins and chlorocruorins), *Comp Biochem Physiol B*, **82**(1): 1-15,

– Zal, F.; Green, B.N.; Lallier, F.H.; Vinogradov, S.N. and Toulmond, A. (1997) *Eur. J. Biochem.*, **243**, 85-92.